# 日本国特許庁 PATENT OFFICE JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 Date of Application:

2000年12月12日

出 願 番 号 Application Number:

特願2000-383118

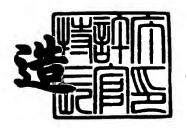
出 願 人 Applicant (s):

株式会社日立製作所

2001年 4月 6日

特 許 庁 長 官 Commissioner, Patent Office





#### 特2000-383118

【書類名】

特許願

【整理番号】

K00014231

【提出日】

平成12年12月12日

【あて先】

特許庁長官殿

【国際特許分類】

G06F 12/00

【請求項の数】

18

【発明者】

【住所又は居所】

神奈川県小田原市中里322番地2号 株式会社日立製

作所 SANソリューション事業部内

【氏名】

小泉 裕

【発明者】

【住所又は居所】

神奈川県小田原市中里322番地2号 株式会社日立製

作所 SANソリューション事業部内

【氏名】

田路 巌

【発明者】

【住所又は居所】 神奈川県小田原市中里322番地2号 株式会社日立製

作所 SANソリューション事業部内

【氏名】

築山 徳広

【特許出願人】

【識別番号】

000005108

【氏名又は名称】

株式会社日立製作所

【代理人】

【識別番号】

100075096

【弁理士】

【氏名又は名称】

作田 康夫

【手数料の表示】

【予納台帳番号】

013088

【納付金額】

21,000円

# 特2000-383118

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

# 【書類名】 明細書

【発明の名称】 ストレージシステムおよびその利用方法

# 【特許請求の範囲】

# 【請求項1】

ストレージシステムの記憶領域ごとに、ストレージ性能に関する要求値の設定 を受けつける設定手段と、

前記ストレージシステムの稼動状況に関する値と、設定した前記要求値とを 比較する比較手段と、

前記比較手段の出力に基づき、前記稼動状況に関する値が前記要求値を満足 しない前記記憶領域を特定する特定手段と、

前記特定手段により特定された記憶領域内に格納されているデータの全部または一部を、他の記憶領域に移動する移動手段とを備えたストレージシステム。

# 【請求項2】

請求項1記載のストレージシステムであってさらに、

前記ストレージシステムの稼動状況に関する値の、単位期間の平均値を算出 する算出手段と、

算出された前記平均値と、前記要求値とを比較する第2の比較手段と、

前記第2の比較手段による比較の結果に基づき、前記単位期間の平均値が前 記要求値を超えている前記記憶領域を特定する第2の特定手段を有するストレー ジシステム。

#### 【請求項3】

データセンタにおけるストレージサービス提供方法であって、

ストレージ性能に関するサービスレベル保証契約を締結するステップと、

前記サービスレベル保証契約にもとづき、ストレージ性能に関する要求値を 設定するステップと、

前記ストレージシステムの稼動状況を監視するステップと、

前記監視の結果、前記稼動状況に関する値が前記要求値を超えた場合に、前 記ストレージサブシステム内のデータ配置を変更するステップ

からなる、ストレージサービス提供方法。

# 【請求項4】

請求項3記載のストレージサービス提供方法であってさらに、

前記ストレージシステムの稼動状況に関する値の、単位期間の平均値を算出 するステップと、

前記平均値が、前記要求値を満足していない記憶領域の契約者に対して、サービスレベル保証契約にもとづいて支払われた利用料の全部または一部を払い戻すステップからなる、

ストレージサービス提供方法。

# 【請求項5】

請求項3または4記載のストレージサービス提供方法であってさらに、

前記ストレージシステムの稼動状況に関する値を契約者に報告するステップ からなる、

ストレージサービス提供方法。

# 【請求項6】

データセンタにおけるストレージサービス提供方法であって、

ストレージシステムの記憶領域ごとに、ストレージ性能に関する要求値の設 定を受けつけるステップと、

前記ストレージシステムの稼動状況を監視するステップと、

前記ストレージシステムの稼動状況が前記設定された要求値を満足しない前 記記憶領域に格納されたデータの全部または一部を、他の記憶領域に移動するス テップ

からなる、ストレージサービス提供方法。

#### 【請求項7】

請求項6記載のストレージサービス提供方法であって、

前記要求値は前記記憶領域ごとに異なる値を有し、データセンタの利用料は 、前記要求値にもとづいて決定される、ストレージサービスの提供方法。

#### 【請求項8】

請求項6記載のストレージサービス提供方法であってさらに、

前記ストレージシステムの稼動状況に関する値の、単位期間の平均値を算出

するステップと、

前記平均値が、前記要求値を満足していない記憶領域について、利用料払い 戻しの対象であることを出力するステップからなる、

ストレージサービス提供方法。

#### 【請求項9】

複数の記憶装置と、その制御装置からなるストレージシステムの、前記複数の記憶装置の記憶領域を利用者に割り当てる方法であって、前記制御装置は、前記割り当てた記憶領域へのアクセス性能に関する要求を受け付け、前記割り当てた記憶領域へのアクセス状況を監視し、前記アクセス状況が前記要求を満足しない場合は前記要求を満足するように前記記憶領域を再割り当てする、ストレージシステムの利用方法。

#### 【請求項10】

請求項9において、前記要求内容に応じて課金し、前記要求を満足しない時間 に応じて前記課金額を減じる、ストレージシステムの利用方法。

#### 【請求項11】

請求項1、2に記載のストレージシステムであって、前記ストレージ性能は、 I/0アクセス可能回数、データ転送量、ディスク空き容量率、ディスク利用率、 データ転送速度、キャッシュ常駐量、のうちひとつ以上から決定されるストレージシステム。

#### 【請求項12】

請求項1における、前記データの他の記憶領域への移動は、キャッシュへのステージング、ミラーディスクの作成、データの冗長度の変更、他の物理ボリュームへのデータ移動のいずれかであるストレージサブシステム。

#### 【請求項13】

請求項3において、前記ストレージ性能は、I/Oアクセス可能回数、データ転送量、ディスク空き容量率、ディスク利用率、データ転送速度、キャッシュ常駐量、のうちひとつ以上から決定されるストレージサービス提供方法。

#### 【請求項14】

請求項3において、前記データ配置の変更は、キャッシュへのステージング、

ミラーディスクの作成、データの冗長度の変更、他の物理ボリュームへのデータ 移動のいずれかであるストレージサービス提供方法。

# 【請求項15】

請求項6において、前記データの他の記憶領域への移動は、キャッシュへのステージング、ミラーディスクの作成、データの冗長度の変更、他の物理ボリュームへのデータ移動のいずれかであるストレージサービス提供方法。

# 【請求項16】

請求項9において、前記アクセス性能は、I/Oアクセス可能回数、データ転送 量、ディスク空き容量率、ディスク利用率、データ転送速度、キャッシュ常駐量 、のうちひとつ以上から決定される、ストレージシステムの利用方法。

# 【請求項17】

請求項9において、前記記憶領域の再割り当ては、キャッシュへのステージング、ミラーディスクの作成、データの冗長度の変更、他の物理ボリュームへのデータ移動のいずれかである、ストレージシステムの利用方法。

#### 【請求項18】

ネットワークに接続されたサーバ、前記サーバに接続された記憶装置からなり、前記ネットワークを介したアクセスを受ける記憶システム群の運用方法であって、

顧客の希望する記憶装置性能についての要求値を、前記記憶装置に接続された記憶装置管理端末が前記記憶装置の記憶領域ごとに受け付けるステップと、

前記要求値と、前記記憶装置の稼動状況を表す数値との関連性を調べるステップと、

前記稼動状況を示す数値が、前記要求値を満足しない前記記憶領域に格納されたデータを、他の記憶領域に移動するステップ

を有する、記憶システム群の運用方法。

# 【発明の詳細な説明】

[0001]

### 【発明の属する技術分野】

本発明は、データを記憶するストレージシステム、およびストレージシステム

の利用方法に関する。

[0002]

# 【従来の技術】

産業のあらゆる分野でIT化が推進される昨今、これまでデータが電子管理されていなかったような業種においても、サーバやストレージシステムの導入によるデータの電子管理が必要となってきている。あるいは、従前よりストレージシステムを活用してデータを電子管理している業種であっても、管理すべきデータは急増している。このようなデータの急増に伴い、業務遂行上必要となるストレージ容量は増加の一途を辿っている。

[0003]

上記のような場合において、自社内に新しくサーバやストレージシステムを導入して管理したり、データの急増に対して業務に支障が生じない的確なタイミングでストレージ容量を増設することは企業にとって負担が大きい。かかる事情に鑑み、データ管理のアウトソーシングを受託し、サーバやストレージをレンタルして利用料を徴収するビジネスが普及しつつある(データセンタ業務)。

[0004]

このようなデータ管理のアウトソーシングビジネスにおける運用方法の例として、ストレージ容量を貸し出し、利用料を徴収する特表2000-501528では、データ記憶装置へのアクセス速度を基に、データ記憶装置を高速、中速、低速に分類して階層化したものが開示されている。本従来技術のストレージの課金方法は、データ記憶装置へのアクセス速度が速い装置ほど単位記憶容量あたりの利用料を高額とするもの、すなわち利用している記憶容量だけでなく、利用しているデータ記憶装置の種類をも利用料金を決定するための要素とするものである。そして、ストレージ利用料の徴収にあたって、課金要素に関するデータをストレージから出力し、高速記憶装置、中速記憶装置、低速記憶装置のそれぞれの利用料を合計して、単位期間ごとの利用料を徴収する方法が開示されている。

[0005]

【発明が解決しようとする課題】

本従来技術によれば、契約により顧客毎にデータ記憶装置は固定的に割り当て

られ、一旦データの格納先であるデータ記憶装置が割り当てられると、データは そのデータ記憶装置に留まる。

[0006]

ところが、ストレージシステムを利用するにあたり、定期的あるいは急激にトラフィックが増加すると、当該ストレージシステム全体又は特定の論理デバイスに対するストレージ性能が低下するという問題がある。ストレージ性能低下は割当てられている記憶媒体の容量の大小とは無関係に起こる。たとえば、空き容量が十分に確保されていても、特定のデータにアクセスが集中すれば、データアクセスが長時間待たされるケースが発生し得る。

[0007]

本発明の目的は、ストレージシステムの利用中に、ストレージシステムの性能をある一定水準に維持するストレージシステムの運用方法を提供することにある

[0008]

また、本発明の目的は、ストレージシステムの性能を設定するのに用いる入力 手段を提供することにある。

[0009]

【課題を解決するための手段】

上記課題を解決するため、サービスレベル保証契約により、顧客ごとにストレージ性能に関する一定のサービスレベルを保証する。本発明においては、ストレージシステムに、ストレージシステムの稼動状況を監視するための性能監視機構と、データ移動手段を導入する。

[0010]

性能監視機構は、デバイスの利用率、データ転送速度等の各種パラメータの関値を設定する機構と、ストレージシステムの運用状況に伴って推移する当該パラメータのモニタリング機構から成る。パラメータをモニタリングした結果、ストレージシステム全体又は特定の論理デバイスにおいてストレージ性能の低下が検出された場合は、データ移動手段によりデータを移動することで、負荷分散を図る。

#### [0011]

# 【発明の実施の形態】

以下、本発明の実施の形態を図面を参照しながら詳細に説明する。

[0012]

図2は、本発明の一実施形態によるデータセンタ(240)と、データセンタ(240)にアクセスしてくるクライアントPCにより構成されるネットワークシステムの構成図である。この図において、LAN/WAN(ローカルエリアネットワーク/ワイドエリアネットワーク・204)よりも下に描かれている構成要素がデータセンタ(240)を構成している。クライアントPC(201~203)はLAN/WAN(204)を経由して、データセンタ(240)にアクセスすることにより、データセンタ(240)と契約しているプロバイダA~C(233~235)の提供する種々のサービスを受けることができる。データセンタ(240)はサーバ(205~207)とその配下のストレージ(209)より構成される。なお、サーバ(205~207)とストレージ(209)は、ストレージエリアネットワーク(SAN・208)で接続されている。

### [0013]

ストレージ (209) の内部構成の一部を、図3に詳細に示した。ストレージ (20 9) 内部には複数種類の記憶媒体が格納されている。(この図では、便宜的にタイプA、タイプB、タイプCの3種類に分かれるとしているが、記憶媒体の種類は3種に限定されない。)各階層の記憶媒体は、ディスクコントローラ (327・図2では227に相当) によって制御され、ディスクコントローラ327はDKA (ディスクアダプタ・317~319) 及びバス/スイッチ (320)、サーバ (205~207) とのインタフェースであるチャネルコントローラ (321)、及びサーバ (205~207) からの入出力に関連するデータを一時的に保持するキャッシュ (322) を有している。その他のストレージの構成要素については、本図では省略する。

#### [0014]

ストレージは、上記各構成要素の性能を監視したり、ストレージの各種動作の 条件設定や起動を制御するプロセッサであるサービスプロセッサ S V P (325)を 有し、性能監視用PC (323) が接続されている。

# [0015]

本発明は、このSVP (325) 上で動作するプログラムである性能監視機構 (324) を利用することで、前述した性能保証を実現する。具体的には、この性能監視機構 (324) は、各構成要素の性能を定量的に示すパラメータを収集し、この収集したパラメータの値と、閾値 (326) を比較する。閾値 (326) は、ストレージのSVP (325) 内に設定される。閾値との比較の結果によっては、性能維持のための処理を開始する。これについては、後で、サービスレベル協定の説明と共に詳細に説明する。なお、閾値との比較は、単純に数値同士の比較のみならず、関数との比較など、フレキシブルな条件比較を含む。

#### [0016]

ところで、SVP (325) はストレージ内部に設置される端末であるため、保守員しか利用することができない。従って、ストレージ外部から性能監視機構 (324) と同様の機能を利用したい場合は、性能監視用PCから用いることができる。言いかえれば、本発明を実現するのに性能監視機構の場所は問題ではなく、ストレージの性能を監視し、それを閾値と比較し、その結果に基づいてストレージを制御できれば、本発明を実現できる。

# [0017]

以下、具体的に説明する。まず、性能監視機構(324)により監視するパラメータの例を以下に列挙する。パラメータは例えば、ディスク空き容量率、ディスク利用率、I/Oアクセス可能回数、データ転送量、データ転送速度等がある。それぞれの定義は、ディスク空き容量率…(空き容量÷契約容量)、ディスク利用率…単位時間あたり、記憶媒体(物理ディスクドライブ)がアクセスされて動いている時間の割合、I/Oアクセス可能回数…単位時間あたりのread/write 完了数、データ転送量…1I/Oあたりに転送できるデータ量、データ転送速度…単位時間あたりに転送できるデータ量である。

## [0018]

ストレージの利用中、特定のデバイスに、あるいは特定の時間帯にアクセスが 集中したり、アクセス急増が起こる結果、ストレージ性能が低下することがある 。ストレージ性能の低下は、上記のパラメータ量が閾値を超えたことにより検出 できる。この場合は、必要とされているストレージ性能を維持すべく、特定のデ バイスに集中している負荷を分散する。

[0019]

局所的にアクセスが集中した結果生じる、ストレージ性能の低下に際しては、 アクセスの分散を行ってストレージ性能を維持する必要がある。

[0020]

本発明では、ストレージにおいてデータの格納場所を分散する方法について述べる。

[0021]

図2に示されるネットワークシステムにおいては、ストレージ (209) やサーバ (205~207) を所有するデータセンタ (240) はプロバイダ (233~235) と、ストレージの記憶容量および特定のサーバを提供することを契約する。プロバイダ (233~235) は自社の契約範囲内の記憶容量を活用して、種々のサービスをエンドユーザのクライアントPC (201~203) に対してLAN/WAN経由で提供する。つまり、本ネットワークシステムは3者の契約 (データセンタ⇔プロバイダの契約、およびプロバイダ⇔エンドユーザの契約の計2契約) により成り立っている。

[0022]

図2はまた、ストレージ、サーバを所有するデータセンタ(240)と、プロバイダ(233~235)、クライアントPC(201~203)の関係の概念図をも示している。エンドユーザは、クライアントPC(201~203)を利用してネットワーク経由でデータセンタ(240)にアクセスする。データセンタ(240)には、エンドユーザが契約しているプロバイダ(233~235)の業務データが格納されており、プロバイダ(233~235)は業務データの管理をデータセンタ(240)に委託し、データセンタ(240)はプロバイダ(233~235)から利用料を回収する。プロバイダが提供するサービスを利用するクライアントは、そのサービスに対する対価を支払う。

[0023]

上記のように、プロバイダは、その設備利用に関してデータセンタと契約する。データセンタから提供されるハード設備の性能(ストレージ、サーバ等の性

能)はプロバイダがクライアントに提供できるサービスの性能に直接関係する。 そのため、ストレージ性能の維持がデータセンタ⇔プロバイダ間の契約の中で保 証されれば、プロバイダはエンドユーザに対して性能的に安定したサービスを提 供できることになる。本発明はこの性能的に安定したサービスを可能とするもの である。

[0024]

本ネットワークシステムを用いたデータセンタ業務において、ストレージ(209)が提供できるストレージ性能を定量化し、提供可能なサービスの内容を透明化するため、サービスレベル協定(SLA)という概念を導入する。

[0025]

ここでサービスレベル協定 (SLA) について簡単に述べる。サービス契約においては、提供されるサービスを定量化し、サービスの質について上限あるいは下限を示して明確にすることが望ましい。これは、サービスの提供を受ける側にとっては、他社のサービスとの比較を容易にするメリットがある。また、自らの希望に適合した適正なサービス内容を適正な価格で受けられるというメリットもある。サービスを提供する側にとっても、提供できるサービスの上限と下限を示すこと、およびサービス提供者の責任範囲を明確にすることで、サービスを受ける顧客から過度な期待を受ける可能性を低減し、トラブル発生時の無用な争いを避けることができるというメリットがある。

[0026]

本発明のサービスレベル協定(SLA)は、上記のデータセンタ、プロバイダ、エンドユーザの3者の契約のうち、データセンタとプロバイダ(233~235)が締結する契約に関する。サービスレベル協定は、上述した性能監視機構(324)により監視する複数のパラメータと、プロバイダが希望する記憶装置の契約容量(ディスク容量)に基づいて決定する。

[0027]

これらのパラメータを用いてデータセンタとプロバイダがサービスレベル協定 を締結する際のフローを説明する。 [0028]

まずデータセンタがプロバイダに保証する保証内容(目標性能)を決めるフローを図4を用いて説明する(サービスレベル協定締結時のフロー:ステップ401~ステップ407)。

[0029]

図4において、プロバイダは、データセンタに保証して貰いたいストレージに関する保証項目、例えば、RAIDグループ単位に保証されるディスク利用率(単位時間あたり、記憶媒体がアクセスされて動いている時間の割合)や、ストレージ容量の空き容量率(空き容量・契約容量)、等の項目から一つを選択する(ステップ402)。この項目の設定に関しては、図5を用いて後で説明する。

[0030]

次に、プロバイダは、この選択した保証項目のそれぞれに対して、保証内容およびそのための値(性能値)を設定する(ステップ403)。例えば、ステップ402で選択した保証項目がドライブのビジー率だったら、「RAIDグループあたりのディスク利用率を平均60%以下に押さえること」、または、「RAIDグループあたりのディスク利用率を平均80%以下に押さえること」等、ディスク利用率の値を指定する。ステップ402で選択した保証項目がストレージの空き容量率だったら、「ストレージの空き容量率は常に20%が確保されるよう増設すること(すなわち、全契約容量に対しての空き容量が20%を切ったら、増設して空き容量を増やすということである。あるプロバイダの契約容量が50ギガバイトであるなら、常に10ギガバイトは未使用の領域として確保されていなければならない。)」、等である。ここで、この「60%」や「80%」が目標性能(保証内容)である。

[0031]

保証項目、及び保証内容が決まると、データセンタは、この内容に関した料金をプロバイダへ提示する。プロバイダは、この提示された料金に対して合意するか、否かを判断する(ステップ404)。このステップは、保証内容が含む保証値により、データセンタがこれら保証内容を実現するのに必要なハードウェア資源の使用量がことなり、結果としてプロバイダへ提示する料金が変わるので、こ

の料金の変動をプロバイダに認証させるものである。また、料金に満足できないプロバイダは、この合意に否と回答を返すことで、再度、保証内容を再設定できるステップへもどることができるので、プロバイダは予算の管理も容易になる。ステップ403及びステップ404に関しては、図6を用いて後で更に説明する

次に、全保証項目に対して保証内容が設定されたかを確認する(ステップ405)。この確認が終了すると、データセンタは、契約した内容の項目を再度出力し、プロバイダに保証項目、保証内容(性能値)、料金等を確認させる(ステップ406)。この際、料金は全保証内容の合計金額も合わせて確認させるほうが望ましい。

[0032]

図5は、図4のステップ402を詳細に説明する図である。図5に示すように、保証内容は、例えば、パソコン画面上にリストの形で表示させることにしておき、これをデータセンタの顧客であるプロバイダに、画面上で選択させることによって実現させる。これにより、プロバイダは保証内容を容易に選択することが可能となる。なお、プロバイダが既に必要な項目を選択済みの場合は、例えば、図には示さないが、図4のステップ402からステップ406へのフローがあると望ましい。

[0033]

図6は、図4のステップ403、ステップ404の一手段を示す図である。図6では、画面上に、プロバイダの業務内容対応にその推奨値および料金が示される。例えば、プロバイダ業務をタイプA(オンライン業務が主で遅延時間に比較的制約が大きい業務)、タイプB(バッチ処理中心で、遅延時間に対する制約が小さい業務)、タイプC(大量データを扱う業務)、等に区分し、それに対するドライブのビジー率の推奨値が例示される。従って、プロバイダは、エンドユーザに対して提供するサービスがどのタイプにあたるかを判断し、そのタイプを選択すれば良い。ここで、この値はあくまでも推奨値であり、その後、プロバイダは、データセンタから報告されるストレージの性能統計データをベースに、これらの値を再設定できるものとする。なお、図6に示す方法は、一例であり、ステ

ップ403、ステップ404に関し、単純に保証レベルである数値を直接入力し 、それに対応する料金を確認する方式でも良い。

[0034]

以上、図4から図6を用いて、サービス保証項目およびその内容の決定方法を示した。これら決定されたサービス保証項目およびその内容は、SVPが有する入力手段を介して、SVPが有するメモリ等の記憶手段に記憶され、別途収集される性能測定値と比較される。そして、その結果に応じてストレージを制御される。なお、SVPへの前記サービス項目やその内容性能目標値の入力は、図4のステップをサポートするパソコンから通信回線を介して入力するようにすれば、SVPの入力手段を使用した入力は不要となる。

[0035]

図4にサービスレベル協定締結時のフローを、図5、図6にサービスレベル選択時にプロバイダが操作する画面を示した。図5の項目選択画面は、図4のステップ402に対応し、図6の閾値設定画面は図4のステップ403に対応している。

[0036]

サービスレベル協定設定は、以下のステップにより実行される。データセンタとの契約を希望するプロバイダは、図5に示した項目選択画面から、設定項目の中から1を選択してチェックボックスをオンにする(ステップ402)。選択した項目の閾値設定画面(図6)が表示され、プロバイダは自己の業務規模、データの種類、予算等に応じて、最適な選択肢を選ぶ。閾値の設定は、図6のチェックボックスをオンにすることで行う(ステップ403)。

[0037]

次に、上記のプロセスで締結されたサービスレベル協定が実際に遵守されるための、データセンタの運用方法について説明する。

[0038]

図7に例として、利用率の監視画面を示した。利用率は、RAIDグループ(後述)単位で保証されている。利用率の監視画面は、SVP(325)、もしくは性能監視用PC(323)から参照することができ、各ボリュームの利用状況を数値で表示する。利用率監視画面は、論理ボリューム番号(701)、当該論理ボリューム

の平均利用率(702)、当該論理ボリュームの最大利用率(703)、いくつかの論理ボリュームを分割して格納している複数の物理ディスクドライブの組からなるRAIDグループの番号、RAIDグループ全体の利用率の平均値と最大値(706)、およびRAIDグループの使用状況を表す情報(704、705)から成る。なお、具体的な定義については、図11を用いて後述する。

# [0039]

RAIDグループの使用状況を表す情報(704、705)について詳説する。RAIDグループは、当該ボリュームを含む複数の論理ボリュームを分割して格納している複数の物理ディスクドライブの組で構成されている。図1に、3つのデータディスクで構成される例でRAIDグループの概念図を示した。(構成ディスク数はこれらに限定されない。)本図においては、4つの論理ボリュームV0~V3を格納している3つの物理ディスクドライブD1~D3が、RAIDグループAを形成している。RAIDグループAから、論理ボリュームV0を除いた論理ボリュームV1~V3で新しいRAIDグループA'を再構成することを考える。

#### [0040]

論理ボリュームVOに関するRAIDグループの使用状況を表す情報(704、705)は、RAIDグループAから論理ボリュームVOを除いてRAIDグループA'を再構成した場合の新しいRAIDグループ全体の利用率を示す情報である。数値は利用率の平均値(704)と最大値(705)がそれぞれ示される。すなわち、論理ボリュームVOを他へ移動した場合に残りの論理ボリュームによる平均ドライブ利用率を示している。

#### [0041]

サービスレベル協定締結後の運用においては、設定したサービスレベル協定に基づいて、閾値を設定し、実際のストレージの利用率(702~705)と閾値の関係を図7の監視画面で常時モニターする。ストレージの利用率を示す数値が設定した閾値を超えることにより、サービスレベル協定で保証されている「平均XX%」等の値を超えそうになった場合は自動、あるいは保守員により手動でデータを移動する。(通常はサービスレベル協定で保証した「平均XX%」を閾値として性能監視機構(324)に設定し、パラメータが閾値を超えた契機でデータを移動する

ことにより、平均値をサービスレベル協定で保証したXX%以下に抑える。) 以下、ドライブの利用率をプロバイダに保証する方法を具体的に説明する。

#### [0042]

まず、図1を用いて、サーバがストレージをアクセスする単位である論理ボリューム(論理デバイス)と実際にデータを記録する物理ドライブの関係を説明する。今、ストレージをRAID(Redundant Array of inexpensive disk)のレベル5構成だと考えると、複数の論理ボリュームは、図1に示すように、複数の物理ドライブ(RAIDグループ)に割り付けられ、この際、それぞれの論理ボリュームが複数の物理ドライブに分散するように割り付けられる。ストレージには、このような複数の物理ドライブから構成されるRAIDグループが多数設けられ、これにサーバからのデータを記録する際の管理単位である論理ボリュームが割り付けられる。なお、RAIDおよびそのレベルに関しては、パターソンの論文『D. Patterson, G. Gibson, and R. Katz, "A Case for Redundant Arrays of Inexpensive Disks (RAID)," Report No. UCB/CSD 87/391 (Berkeley: University of California, December 1987)』に記載がある。また、図1はRAIDグループを物理ドライブD3個により構成しているが、この数は3個に限定されない。

## [0043]

上記のように、複数の物理ドライブに複数の論理ボリュームが割り付けられている場合、或る特定の論理ボリュームにサーバからのアクセスが集中すると、当該特定論理ボリュームが割り付けられているRAIDグループに属する他の論理ボリュームへのアクセスを圧迫することになる。また、RAIDグループに属する複数の論理ボリュームへのアクセスが全体的に増加すると、RAIDグループに属する物理ドライブのビジー率が高くなり、その結果、各論理ボリュームに対するアクセス遅延時間が急激に増加する。そこで、これら各論理ボリュームへのアクセスを監視し、ドライブへのアクセス状況に関する統計データを収集し、ある論理ボリュームを利用率が低い他のRAIDグループへ移動することで、当該RAIDグループのビジー率をある特定値以下に保証できる。

#### [0044]

即ち、プロバイダは、論理ボリュームを使用する際、RAIDグループに属する物理ドライブの利用率を一定値以下にする契約をデータセンタとの間に結んだ場合、データセンタは、ストレージ内のRAIDグループのアクセス状況を監視し、必要に応じてRAIDグループ上の論理ボリュームを他のRAIDグループへ移動することで、その性能値をプロバイダに対して保証することができる。

#### [0045]

図11は、RAIDグループ1の性能を管理する性能管理テーブルの一例を示 す。この性能管理テーブルは、ストレージ内のRAIDグループ対応に設けられ 、SVP内の性能監視機構で管理される。本テーブルには、まず、本RAIDグ ループ1に属する各ドライブ(D1,D2、D3)対応に、各ドライブに割り付 けられた各論理ボリューム(VO、V1、V2、、、)毎に、各ドライブに対する単位 時間あたりのアクセス時間として、利用率が示される。例えば、図11のドライ ブD1に対しては、論理ボリュームV0の利用率が15%(単位時間である100 秒の間に15秒のアクセスがドライブD1の論理ボリュームV0へあるということ )、論理ボリュームV1の利用率が30%(単位時間である100秒の間に30 秒のアクセスがドライブD1の論理ボリュームV1へあるということ)、論理ボリ ユームV2の利用率が10%(単位時間である100秒の間に10秒のアクセス がドライブD1の論理ボリュームV2へあるということ)となる。従って、ドライ ブD1の利用率(単位時間あたりの各論理ボリュームへの合計値)は、55%と なる。同様に、ドライブD2での利用率は、論理ボリュームV0は10%、論理ボ リュームV1は20%、論理ボリュームV3は10%、ドライブD2の利用率は4 0%となる。同様に、ドライブD3での利用率は、論理ボリュームV0は7%、論 理ボリュームV1は35%、論理ボリュームV2は15%、ドライブD2の利用率 は57%となる。従って、RAIDグループに属するドライブの平均利用率は、 3つのドライブの利用率の平均であり、50.7%となる。また、RAIDグル ープ内のドライブの最高利用率はドライブD3の57%となる。

[0046]

図12は、他のRAIDグループ2に論理ボリュームV3と論理ボリュームV4

が割りつけられている場合を示す。この場合、ドライブD1の利用率は15%、ドライブD2の利用率は15%、ドライブD3の利用率は10%、RAIDグループに属するドライブの平均利用率は13.3%となる。

#### [0047]

なお、これらドライブの利用率はディスク制御装置DKCのDKAが、ドライブをアクセスしてから応答を受けるまでの時間をドライブアクセス時間として求め、性能監視機構に報告することができるが、ディスクドライブ自体が各論理ボリュームへのアクセスを区別することができれば、ディスクドライブ自体がこれらのアクセス時間を計り、これを性能監視機構へ報告するようにしても良い。なお、このドライブ利用率の計測は、そのシステムでの定義に従って矛盾無く計測されれば良いものであり、ドライブの利用状態を客観的に一定の条件で表現できるものであれば、その定義は自由である。

#### [0048]

今、データセンタがプロバイダに保証しているドライブの利用率をRAIDグループに属するドライブの平均利用率が60%以下を保証するとしたとする。一般に、システムがある動作するには遅延が発生するので、あるRAIDグループのドライブの平均利用率を60%以下とした場合、それより低い値の利用率(閾値)で動作を開始する必要がある。本実施例では、契約での利用率の保証値を60%以下とし、その性能を保証するための動作を開始する利用率(閾値)を50%とする。

#### [0049]

先に説明した図11では、RAIDグループのドライブの平均の平均利用率が50%を越えており、このままでは、RAIDグループ1のドライブの平均利用率が60%を越える可能性がある。そこで、SVPは性能監視機構は、RAIDグループ1から論理ボリュームの一つを他のRAIDグループへ移動することで、当該RAIDグループのドライブの平均利用率を50%以下へ下げる動作を開始する。

#### [0050]

この際、この動作を開始するに当たり2つのことを決定する必要が有る。一つ

は、RAIDグループ1からどの論理ボリュームを他のRAIDグループへ移動 するか、もう一つはどのRAIDグループへ移動するかである。

[0051]

RAIDグループ1からどの論理ボリュームを移動させるかを決定するのに、 移動元であるRAIDグループ1の平均利用率が50%以下となるような論理ボ リュームを選ぶことが必須である。図11では、当該ボリュームを他へ移動した 後のRAIDグループ1のドライブの平均利用率も併せて示す。この場合、論理 ボリュームVOを他へ移動した場合にその残りの論理ボリュームによる平均ドラ イブ利用率は40%、(図1のRAIDグループAからA'への再構成に相当)論理ボ リュームV1を他へ移動した場合にその残りの論理ボリュームによる平均ドライ ブ使用率は22.3%、論理ボリュームV2を他へ移動した場合にその残りの論 理ボリュームによる平均ドライブ使用率は39.0%であり、何れも50%以下 を満足する。従って、どの論理ボリュームを選んでも良い。本実施例では、RA IDグループ1の平均利用率が最も小さくなる論理ボリュームV2を移動するも のとして説明をする。なお、論理ボリュームを選択する際、平均利用率が50% 以下という条件のほかに、移動する論理ボリュームへのアクセスが少ない論理ボ リュームを選択する方がそのアクセスに影響を与えないことから、例えば、図1 1の場合、論理ボリュームへの平均利用率がもっとも少ない論理ボリュームV0 を選んでも良い。また、図には示さないが、論理ボリュームに含まれる実際のデ ータ量が少ないほど論理ボリュームを移動するための時間が少なくてすむわけで あるから、論理ボリューム毎のデータ量を管理しておき、最もデータ量の少ない 論理ボリュームを選択するようにしても良い。

[0052]

次に選択した論理ボリュームをどこへ移動するかを決定する必要がある。これを決定するには、現在の平均ドライブ利用率が50%以下であること、かつ、選択した論理ボリュームを移動した先のRAIDグループが、選択した論理ボリュームを移動された後でも平均ドライブの利用率が50%(閾値)以下であることが、プロバイダとの契約を守るため必要である。図13は、RAIDグループ2にRAIDグループ1の論理ボリュームV1を移動する場合の予測表を示してい

る。RAIDグループ2はドライブの平均利用率が現状13.3%であり、他RAIDグループから論ボリュームを受け入れる余裕がある。この状態で、論理ボリュームV1を受け入れた状態が予測として示される(図13の下部)。図示されるように、移動を受け入れた後のドライブの平均利用率の予測値は41.7%であり、閾値より小さく、受け入れ可能と判断することができる。この後、論理ボリュームV1のRAIDグループ1から2への移動を正式に決定する。このように、性能を保証するためには、移動元のRAIDグループの利用率を保証するはか、移動先のRAIDグループの利用率を計算、予測し、利用率を保証できることを確認してから移動することが必要である。予測値が利用率50%を越えた場合には、他のRAIDグループのテーブルを参照し、前述と同様の処理を行う

# [0053]

以上説明したように、論理ボリュームの移動元と移動先の両RAIDグループ において、データセンタはプロバイダに約束した保証値を提供することができる

## [0054]

なお、上記例では、移動の判定条件として閾値50%、また受け入れの条件として閾値50%を用いたが、移動判定条件と受け入れ判定条件が同じ値であると移動に続く移動が発生するおそれがあるため、移動の判定条件の閾値より受け入れの閾値を低く設定する方が望ましい。

### [0055]

また、RAIDグループ内のドライブの利用率を示すのに、前述した平均利用率を用いたが、RAIDグループへのアクセスは最高利用率のドライブに全アクセスのレスポンスが影響を受けるため、最高利用率のドライブに対して、プロバイダとデータセンタとの間で保証値、およびそれに基づく閾値を決定して、これを保証するようにしても良い。

### [0056]

更に、上記説明では、移動元のRAIDグループ1に属するドライブの性能と 移動先のRAIDグループ2に属するドライブの性能が同等として図13を説明 しているが、移動先のRAIDグループ2に属するドライブの性能が移動元のドライブの性能より高い場合がある。例えば、ドライブに対する読み書きの速度が速いような場合、ドライブに対する利用時間が短くなるわけであり、このような場合、RAIDグループ2での論理ボリューム受け入れ後の利用率を計算する際、その性能差を反映させる係数をRAIDグループ1での論理ボリュームV1の各ドライブに対する利用率に乗じたものを、RAIDグループ2の各ドライブの利用率に加えるようにすると良い。性能が悪いドライブへ移動する場合は、逆の係数を用いることで対応できる。

#### [0057]

上記処理は、性能管理機構(ソフトウェア)がスケジューラに従って定期的に チェックし閾値を越えた場合自動的に実行するようにしても良いが、保守員が上 記性能状況テーブルや、予測値テーブルを参照することで、論理モジュールの移 動が必要か判断し、必要と判断した場合に論理モジュールを移動する指示をスト レージに対して入力するようにしても良い。

#### [0058]

なお、上記例ではストレージ内に属するRAIDグループは同じ保証値を有するとして説明しているが、図3に示すようにタイプA、タイプB、タイプCとわけ、タイプAは保証値40%、タイプBは保証値60%、タイプCは保証値80%という様に性能で分ける様にしても良い。この場合、各性能内のRAIDグループ間で論理ボリュームが移動されることになる。

#### [0059]

以上、サービスレベル協定による性能保証内容の決定手順、及び、物理ディスクドライブの利用率を例にその性能を保証する例を示した。 次に、図8では、実際の運用でのサービスレベル協定を実施する上での手順を、データ移動により性能を保証する場合を例に説明する。

# [0060]

業務開始時、あるいは、必要に応じて適宜、サービスレベル協定で保証した閾値に基づいて、性能監視機構324に手動でパラメータの閾値を設定する(ステップ802)。性能監視機構は、監視している実際の機器の性能測定値(パラメ

ータ)が閾値の上限を上回るかあるいは下限を下回る判断する(ステップ803 、ステップ804)。ここでは、閾値を上限値(MAX値)と下限値(MIN値 )の2つを定義しておき、パラメータが上限値を超えた場合は、その性能を保証 できなくなりそうだと、パラメータが下限値より下がりそうだったら、性能に余 裕がありすぎてユーザは過剰スペックな使用をしていると判断する(これに関し ては後述する)。パラメータが、閾値である平均値がXX%を超えた場合は、デ ータを移動することで解決する問題か否かを判定する(ステップ805)。この 判断は、図7、図11~図14を用いて説明したように、移動元および移動先で のRAIDグループに属する物理ドライブの利用率を予測することで行う。スト レージのパフォーマンスを維持できる移動先記憶媒体が存在する場合は、データ を移動する(ステップ807)。このデータ移動は、保守員の判断により手動で 行う場合、サーバのソフトウェアで行う場合、あるいはストレージのマイクロプ ログラムで行う場合がある。すでに、ストレージが提供できるパフォーマンスが 最大限発揮されており、その結果、移動先記憶媒体が存在しない場合は、SVP 325もしくは性能監視用PC323にその旨をメッセージ表示し、必要があれ ばプロバイダに通知する。なお、具体的なデータ移動は、例えば、特開平9-2 74544に開示されているストレージの内部構成、ソフトウェア等を用いるこ とで実現できる。

#### [0061]

図9は、プロバイダに提出するレポートの作成フローを示す。このレポートは、ストレージの稼動状況を含んでおり、定期的にプロバイダに対して報告される。ストレージの稼動状況は、性能監視機構324で監視している各種パラメータにより知ることができる。性能監視機構でデータを収集し(ステップ902)、サービスレベル協定で保証したパフォーマンス(平均XX%以下、等)を満足しているかどうかを判定し(ステップ903)、サービスレベル協定(SLA)を満足している場合は定期的なレポートを作成し、プロバイダに提出する(ステップ904、906)。サービスレベル協定を満足していない場合は、ペナルティ発生レポートを作成し、プロバイダに対して料金割引を行う旨を報告する(ステップ905、906)。



以上、ストレージシステムにおいて、局所的に(特定の物理ドライブに)アクセスが集中して利用率が契約で結んだ性能保証値を越えそうになった場合に、当該物理ドライブに属する論理ボリュームを移動することで、アクセスを分散させることを説明した。上記述べたストレージの局部的な負荷集中を分散させるための他の方法として、負荷が集中しているデータ(図7の例では、利用率が特に高くなっているデータ)のミラーディスクを一時的に作成し、アクセスをそれぞれに振り分けて分散させる方法をとることで性能保証値を保つことも可能である。この場合は、ミラーディスクを作る元のドライブにもアクセスが平均的に半分残る形となる点に注意する必要がある。即ち、当該論理ボリュームの利用率の半分に相当するアクセスは、そのまま現在の物理ドライブに続くことを考慮して、保証動作後の利用率を予測する必要がある。

# [0063]

さらに、上記の負荷分散のための(保証性能値を守るための)データ移動は、同一筐体内の別のRAIDグループを構成する物理ドライブにデータ(論理ボリューム)を移動することを前提としているが、図10に示すように、同じストレージエリアネットワーク(SAN)に接続されている別の筐体へのデータ移動でも良い。この場合は、「小容量であるが高速な記憶装置を搭載した筐体」、「大容量であるが低速な記憶装置を搭載した筐体」のように筐体ごとに保証するパフォーマンスを変えておくという運用方法もある。なお、移動先(別筐体)を決定するには、別筐体ということで、他筐体のRAIDグループを構成する複数の物理ドライブの平均利用率等を入手しておき、これを用いて論理ボリューム移動後の移動先での利用率を予測することで行う。これら他筐体の複数の物理ドライブの平均利用率等は、定期的にSAN上でメッセージを交換することで入手したり、必要に応じて問い合わせることで知ることができる。

## [0064]

また、プロバイダとデータセンタ間で締結したサービスレベル協定は適宜見直 しを行い、初期設定時のサービスレベルではパフォーマンス過剰/もしくは不足 が起こっている場合は、サービスレベルを設定しなおして契約更改する。たとえ



ば、図6で、「 XX > YY > ZZ 」とし、あるデータが格納される物理ドライブの利用率をタイプBの利用率平均YY%で契約したとする。この場合に、長期間にわたってこのデータの利用率平均がZZ%を下回っているような場合は、パフォーマンスが過剰となっているので、サービスレベルをタイプCの利用率平均ZZ%に設定し直し、契約を更改する。これにより、データセンタにとっては高パフォーマンスを提供できるストレージ領域を開放して新たな顧客を掴むチャンスが得られ、またプロバイダにとってはコストを削減するというメリットが得られる。

# [0065]

また、サービスレベル協定の一例として、特定のデータについて、特定の期間だけ契約タイプを変更するという運用方法もある。たとえば、平常時は大容量低速の記憶装置に格納してあるデータを、当該コンテンツの新聞広告掲載に合わせて、1ヶ月間だけ少容量高速の記憶装置に移動し、その期間だけ追加利用料を支払う等の利用形態が考えられる。この場合は、データセンタはあらかじめプロバイダから、対象データ、対象期間の連絡を受け、この連絡内容に含まれる目標性能値を用いてその性能を保証する動作を行うようにする。

#### [0066]

上記では、RAIDグループを構成する物理ドライブの利用率を保証することを説明したが、他の性能保証項目、たとえばディスク空き容量率、I/Oアクセス可能回数、データ転送量、データ転送速度等に関しても、その定義された性能を満足するように、ストレージシステムを動作させることで、サービスレベル協定に従ったサービスを提供することができる。

#### [0067]

例えば、全契約容量に対するディスク空き容量の割合を常に20%確保するというサービスレベル協定を結んだとすれば、ストレージシステムをプロバイダへ貸し出しているデータセンタは、プロバイダが契約しているディスク容量と、実際に使用しているディスク容量とを比較し、空き容量が20%を切ったら、前記プロバイダに対する容量を新たに割り付け、常時空き容量が20%確保されるようにする。等の方法で、この項目に関するサービスレベル保証がなされる。

[0068]

ところで、上記実施例および図3においてはサーバとストレージシステム間は ストレージエリアネットワークで接続されているとしているが、本発明のサーバ ⇔ストレージ間の接続はネットワーク接続に限定されない。

. [0069]

【発明の効果】

以上説明したように、本発明によれば、ストレージシステムの稼動状況に応じてデータの格納先を最適化し、局所的に過剰となっている負荷を分散することができる。これにより、急激なトラフィックの増加が起こっても、ストレージシステムの性能を、契約で保証されている一定基準に保つことができる。

【図面の簡単な説明】

【図1】

RAIDグループの概念図である。

【図2】

データセンタ、プロバイダ、クライアントPC(エンドユーザ端末)の関係を表す概念図である。

【図3】

性能監視機構を搭載したストレージシステムの詳細図である。

【図4】

サービスレベル協定(SLA)設定のフロー図である。

【図5】

SLA設定時のユーザインタフェースの一部であるSLA項目選択画面である

【図6】

SLA設定時のユーザインタフェースの一部である閾値設定画面である。

【図7】

ディスク利用率監視画面の一例である。

【図8】

データ移動判定フローである。

# 【図9】

ストレージの稼動状況レポート作成フローである。

【図10】

筐体外へのデータ移動の概念図である。

【図11】

性能監視画面の一例である。

【図12】

性能管理テーブルの一例である。

【図13】

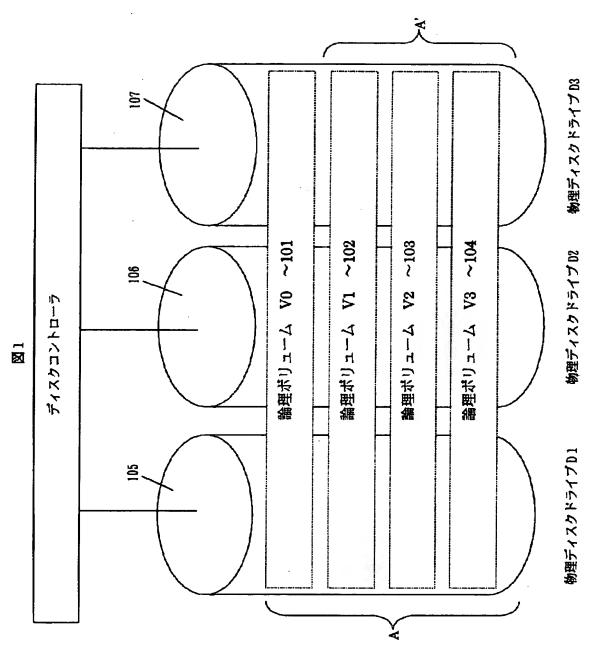
移動後の予測値を含む、性能管理テーブルの一例である。

# 【符号の説明】

- 1101~4…論理ボリューム
- 1105~7…物理ディスクドライブ
- 308…ストレージエリアネットワーク
- 309…ストレージ
- 323 … 性能監視用PC
- 3 2 4 · · · · 性能監視機構
- 325····サービスプロセッサ (SVP)
- 3 2 6 · · · · 閾値
- 327…ディスクコントローラ
- 601~3…サービスレベルタイプ
- 701…論理ボリューム番号
- 702…平均ドライブ利用率
- 704…当該ボリューム移動後の平均ドライブ利用率

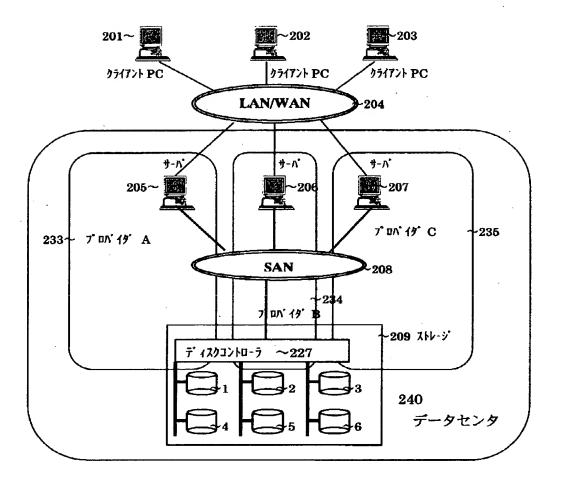
【書類名】 図面

# 【図1】

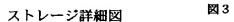


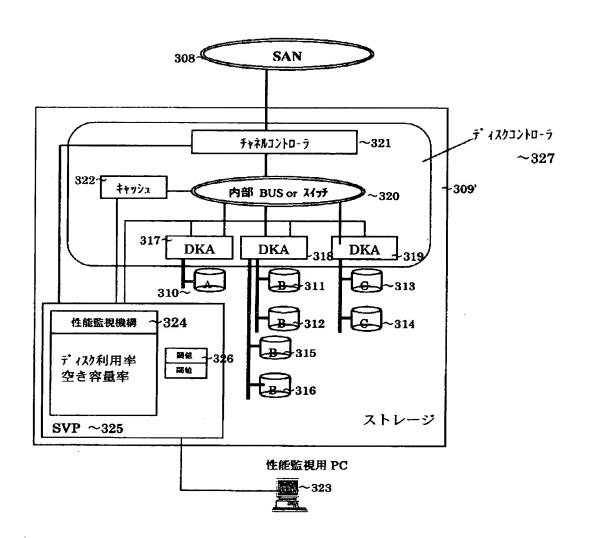
【図2】

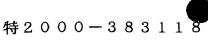
図2

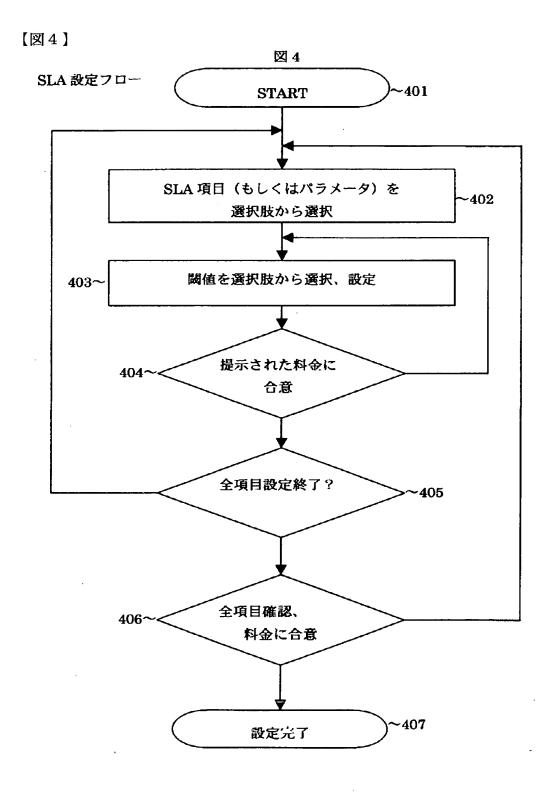


# 【図3】









# 【図5】

# 項目選択画面

## 図 5

501 □ I/O アクセス可能回数 502 □ データ転送量 503 □ ディスク利用率 504 □ データ転送速度 ・

# 【図6】

# 閾値設定画面

図 6

# ディスク利用率

タイプA二平均XX%AA 円/月~601タイプB二平均YY%BB 円/月~602タイプC二平均ZZ%CC 円/月~603

タイプA… 一般ユーザがアクセスしてくるデータに適する

タイプB… 顧客データに適する

タイプ C… アプリケーションファイルに適する

【図7】

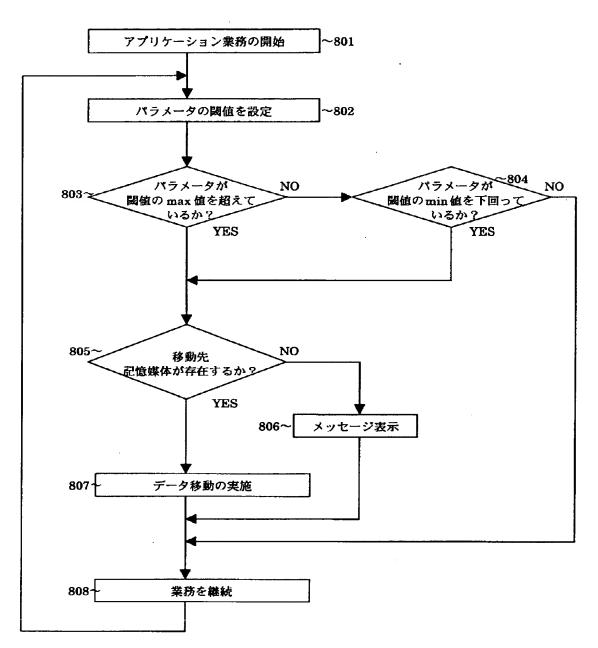
図 7

701	702	703	704	705
論理ボリューム 番号	平均ドライプ	最高ドライプ	平均ドライプ 利用率(移動後)	最高1° 517° 利用率(移動後)
VO	10. 7%	15%	40%	50%
V1	28. 3%	35%	22. 3%	25%
V2	16. 7%	15%	39%	45%
•	•	•	•	•
•	•	•	•	•
•	•	•	•	•
RAID グルー		利田率(設定	値): 平均 609	6 U F

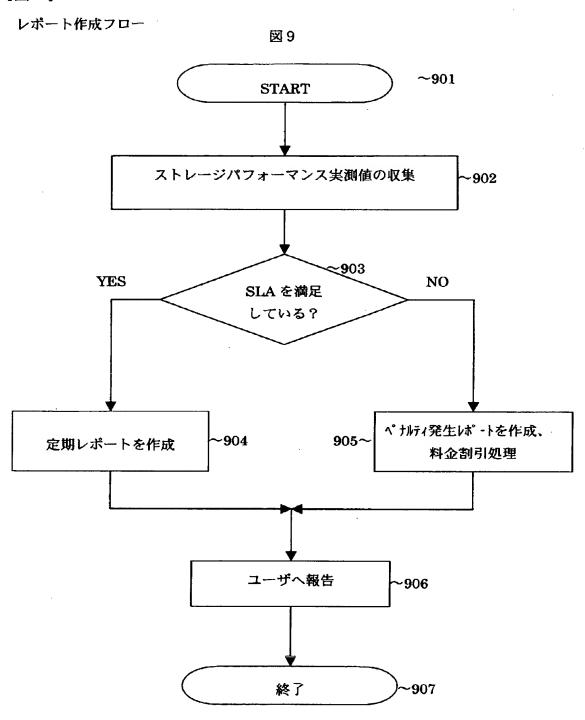
出証特2001-3027794

【図8】

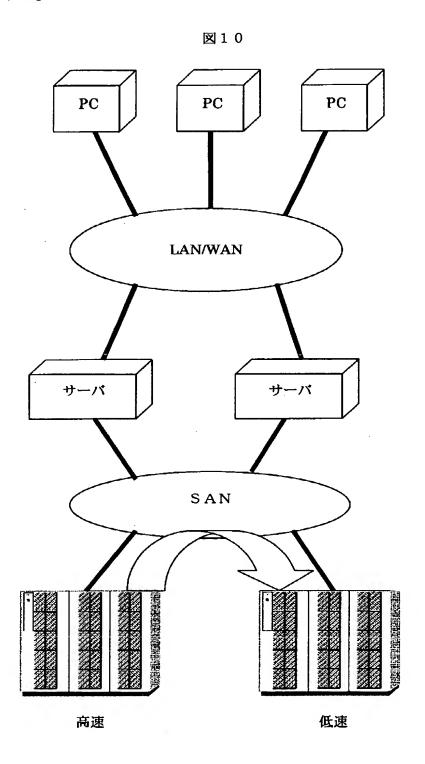




# 【図9】



【図10】 **筐体外** migration



【図11】

		訓 海	
	20.7%	本均	10.7
	平均ドライブ利用率	D3	L
- - -	中地下	D2	10
		D1	15
	RAID JIV-J1		0/0
	RAID 3		

40

28.3

35

20

8

7

33

16.7

9

10

**V**2

50.7

57

40

55

ドライブ 全体

# 【図12】

図12

RAID グループ2

平均ドライブ利用率 13.3%

	D1	D2	D3	平均
V3	10	5	5	6.7
V4	5	10	10	6.7
ドライブ 全体	15	15	10	13.3

【図13】

図13

RAID グループ2

平均ドライブ利用率 13.3%→41.7%(予測)

	D1	D2	D3	平均
V3	10	5	5	6.7
V4	5	10	10	6.7
ドライプ <b>全体</b>	15	15	10	13.3
V1	30	20	35	
ドライブ 全体予測	45	35	45	41.7

# 【書類名】 要約書

【要約】

【課題】

電子化されたデータのアウトソーシングサービスを行うデータセンタにおいて、ある記憶領域に対するアクセスの急増によりストレージシステム全体、または特定のデバイスについてのストレージ性能が低下するという問題がある。

# 【解決手段】

サービスレベル保証契約により、個々の顧客に対して、顧客が選択した一定の サービスレベルを保証する。ストレージ性能が低下しそうな場合は、アクセスが 集中している記憶領域に格納されているデータのデータ移動などを行なってアク セスを分散する。

【選択図】 図3

# 出願人履歴情報

識別番号

[000005108]

1. 変更年月日

1990年 8月31日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台4丁目6番地

氏 名

株式会社日立製作所